Optimizing Emergency Room Wait Times in Ontario's Healthcare System

---

Team Members (Group-7):

- **Dheeraj Kuldeep Choudhary** - 9014533
- **Harika Ravi** - 9032698
- **Nasr Alani** - 9039537
- **Paljeet Singh Sambhi** - 8918495
- **Priyanka Chitikela** - 8909667

Case Study:

Ontario’s emergency rooms (ERs) struggle with excessive wait times, averaging eight hours—twice the global standard. The most significant delay occurs during the initial assessment, where patients first meet healthcare professionals, often taking up to three hours. Addressing this inefficiency is crucial for improving patient care and hospital operations. Traditional solutions, such as increasing staff or expanding facilities, are limited by budget constraints and workforce shortages. Existing research does not fully address how inefficient processes, miscommunication, and technological gaps contribute to these delays.

This study examines the role of patient prioritization, information flow, and resource management in ER efficiency. It explores how AI-driven triage systems and real-time wait-time displays can optimize the initial assessment process and reduce delays. By identifying root causes and evaluating technology-based solutions, this research aims to provide practical recommendations for improving ER workflows and enhancing healthcare service delivery in Ontario.

Import libraries

In [36]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```python
from scipy import stats
import numpy as np
```

Load and display the dataset

In [37]:
```python
# Load the dataset
file_path = "./er_wait_times_ontario_updated.csv"
df = pd.read_csv(file_path)

# Display the first few rows to understand the structure
print(df.head())
```

```
          Hospital    Region        Date  Age               CTAS_Level  \
0     St. Michael's   Toronto  2024-12-14   55  CTAS I-III (Discharged)
1  Hamilton General  Hamilton  2024-03-28   27   CTAS IV-V (Discharged)
2        Sunnybrook   Toronto  2024-10-20   74  CTAS I-III (Discharged)
3     St. Michael's   Toronto  2024-06-18   39  CTAS I-III (Discharged)
4   Toronto General   Toronto  2024-05-10   53  CTAS I-III (Discharged)

   WaitTime_Hours  ER_Occupancy DayOfWeek  Season
0             2.0            84  Saturday  Winter
1             1.4            97  Thursday  Spring
2             7.6            68    Sunday    Fall
3             2.8            73   Tuesday  Summer
4             5.2            98    Friday  Spring
```

Check for missingness in dataset

In [38]:
```python
# Check for missing values
missing_values = df.isnull().sum()

# Calculate percentage of missing values
missing_percentage = (missing_values / len(df)) * 100

# Combine both into a DataFrame
missing_data = pd.DataFrame({
    'Missing Values': missing_values,
    'Percentage (%)': missing_percentage
})
```

```python
# Display the missing values report
print("\nMissing Values Report:")
print(missing_data)

# Check if there are any missing values
if missing_values.any():
    print("\nThere are missing values in the dataset.")
else:
    print("\nNo missing values found in the dataset.")
```

```
Missing Values Report:
                Missing Values  Percentage (%)
Hospital                     0             0.0
Region                       0             0.0
Date                         0             0.0
Age                          0             0.0
CTAS_Level                   0             0.0
WaitTime_Hours               0             0.0
ER_Occupancy                 0             0.0
DayOfWeek                    0             0.0
Season                       0             0.0

No missing values found in the dataset.
```

Exploratory Data Analysis

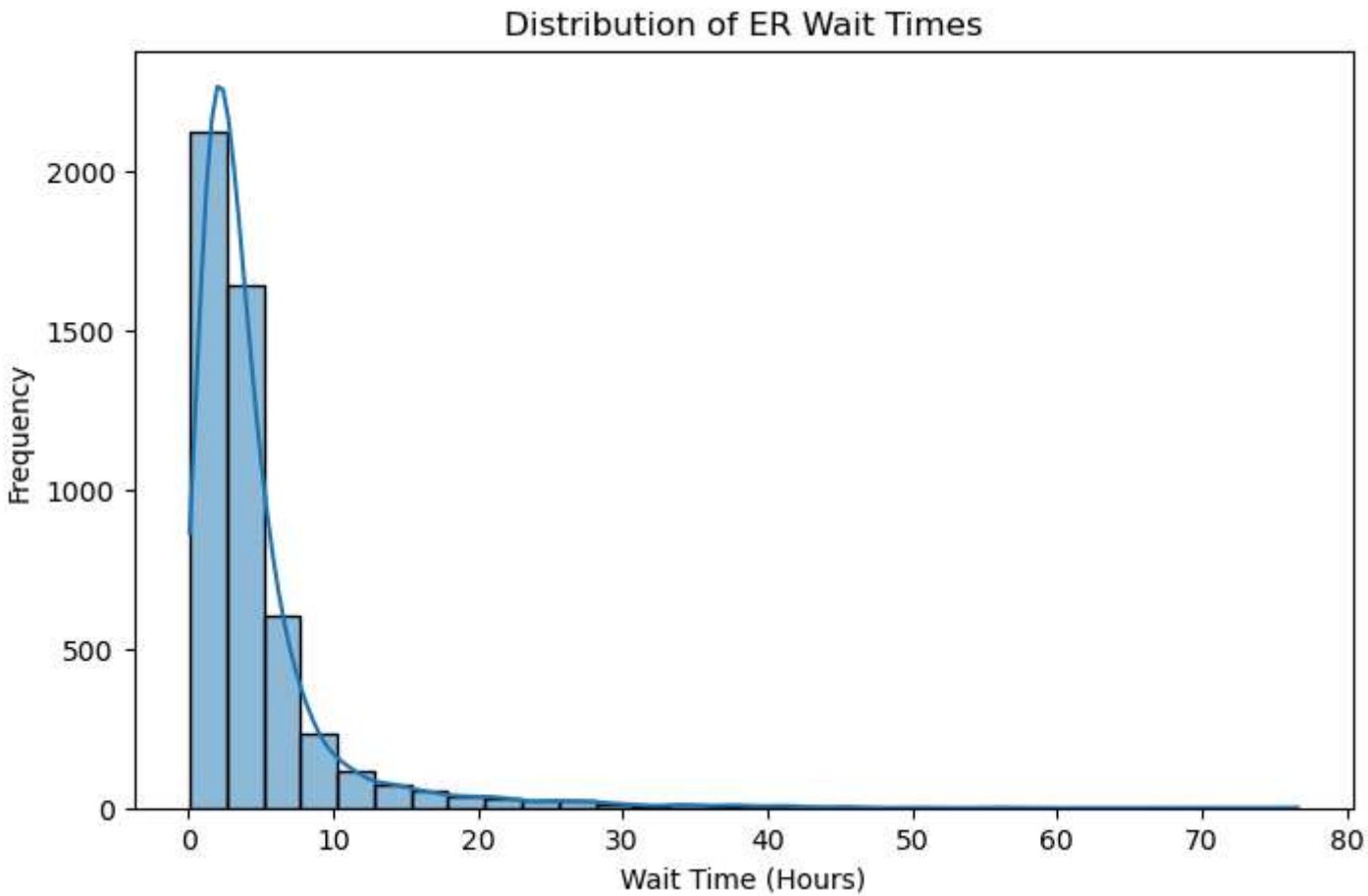
Summary Statastics

In [47]:
```python
# Summary Statistics
summary_stats = df.describe()
display(summary_stats)
```

	Age	WaitTime_Hours	ER_Occupancy
count	5000.000000	5000.000000	5000.000000
mean	44.761200	4.555780	104.026000
std	19.743083	5.338732	55.065733
min	0.000000	0.100000	10.000000
25%	31.000000	1.800000	56.000000
50%	45.000000	3.000000	105.000000
75%	58.000000	5.100000	152.000000
max	100.000000	76.600000	199.000000

Distribution of ER Wait Times (Histogram)

In [40]:
```
# Histogram of Wait Times
plt.figure(figsize=(8, 5))
sns.histplot(df["WaitTime_Hours"], bins=30, kde=True)
plt.xlabel("Wait Time (Hours)")
plt.ylabel("Frequency")
plt.title("Distribution of ER Wait Times")
plt.show()
```

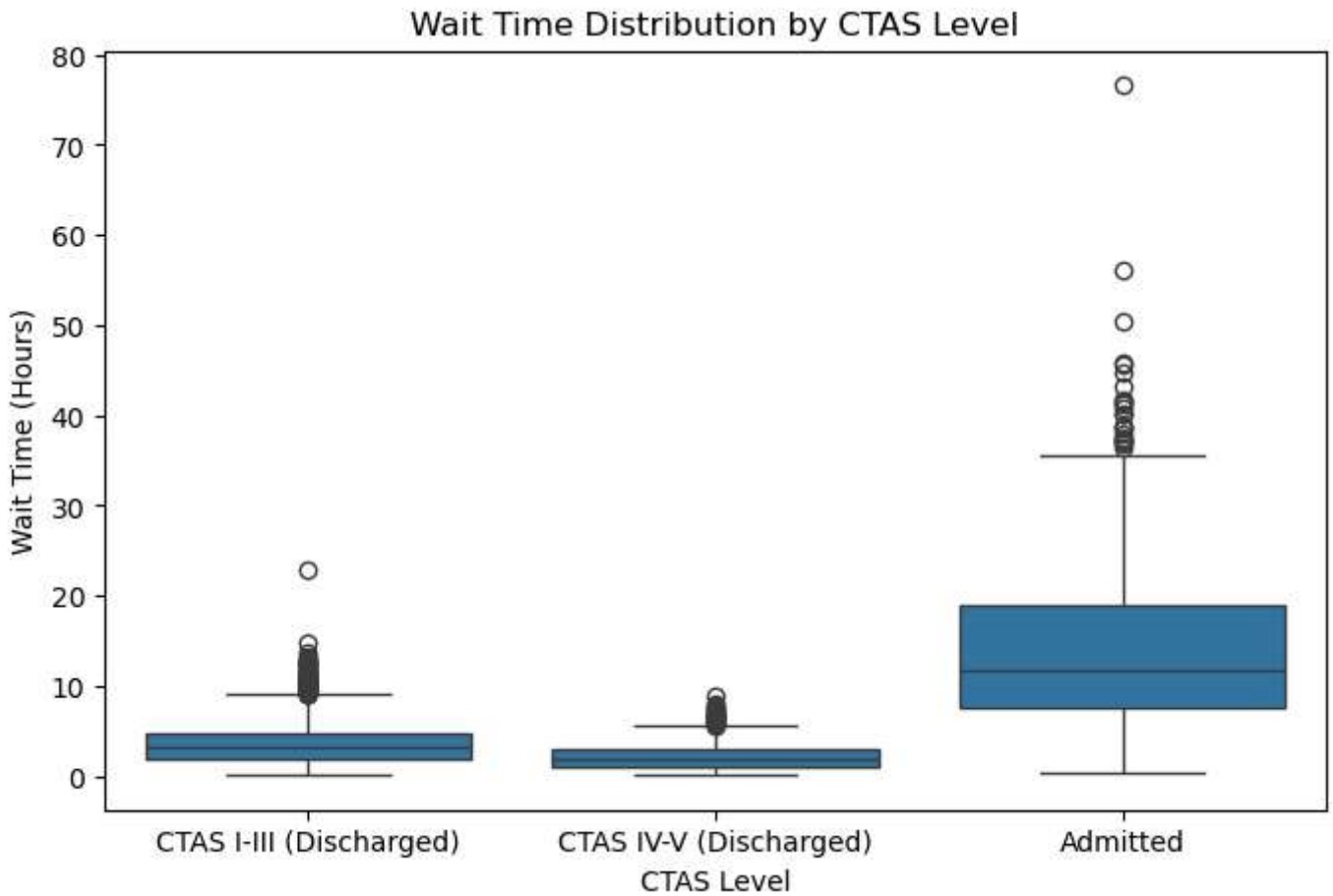


Interpretation:

- The histogram shows that ER wait times are right-skewed, meaning most patients experience shorter wait times, but there are a few cases with extreme delays.
- The highest frequency of ER visits occurs within 0-5 hours, but some patients experience much longer wait times, extending beyond 30+ hours.
- The kde (kernel density estimate) curve further supports the right-skewed distribution, suggesting that most patients do not wait very long, but some extreme cases significantly increase the average wait time.
- This insight highlights the need for better resource allocation and faster triage processes to reduce extreme waiting periods.

Boxplot of ER Wait Times by CTAS Level:

In [41]:
```
# Boxplot of Wait Times by CTAS Level
plt.figure(figsize=(8, 5))
sns.boxplot(x="CTAS_Level", y="WaitTime_Hours", data=df)
plt.xlabel("CTAS Level")
plt.ylabel("Wait Time (Hours)")
plt.title("Wait Time Distribution by CTAS Level")
plt.show()
```



Interpretation:

The boxplot compares wait times across different CTAS (Canadian Triage and Acuity Scale) levels:

- CTAS I-III (Discharged): Patients in this category have moderate median wait times with some outliers beyond 10-20 hours.
- CTAS IV-V (Discharged): These patients generally have shorter wait times compared to CTAS I-III, but still, some experience outliers beyond the expected timeframe.
- Admitted Patients: This category has the longest wait times, with a higher median wait time and a significant number of outliers. Some admitted patients wait more than 40+ hours.

Hypothesis

T-test Statastical Test

Hypothesis:

H_0 (Null Hypothesis): There is no significant difference in average ER wait times between CTAS I-III and CTAS IV-V patients.

H_1 (Alternative Hypothesis): There is a significant difference in average ER wait times between CTAS I-III and CTAS IV-V patients.

In [42]:
```
# Checking if there is a significant difference in wait times between CTAS I-III and CTAS IV-V
ctas_i_iii = df[df["CTAS_Level"] == "CTAS I-III (Discharged)"]["WaitTime_Hours"]
ctas_iv_v = df[df["CTAS_Level"] == "CTAS IV-V (Discharged)"]["WaitTime_Hours"]

t_stat, p_value = stats.ttest_ind(ctas_i_iii, ctas_iv_v, equal_var=False)  # Welch's t-test

# T-Test Results Display
print("\nT-Test Results:\n")
print(f"T-Statistic: {t_stat:.4f}")
print(f"P-Value: {p_value:.4e}")

# Hypothesis Testing
alpha = 0.05  # Significance level

# Interpretation
if p_value < alpha:
    print("\nConclusion: Reject Null Hypothesis H₀, meaning CTAS I-III and CTAS IV-V patients have significantly diff
else:
    print("\nConclusion: Fail to reject Null Hypothesis H₀, meaning there is no significant difference in ER wait tim
```

```
T-Test Results:

T-Statistic: 25.3275
P-Value: 2.1990e-130

Conclusion: Reject Null Hypothesis H₀, meaning CTAS I-III and CTAS IV-V patients have significantly different ER wait times.
```

Interpretation of T-Test Results:

- CTAS (Canadian Triage and Acuity Scale) categorizes ER patients based on urgency, from
  - CTAS I–III (higher urgency) to CTAS IV–V (lower urgency).
- The high T-statistic (25.3275) indicates a strong difference in wait times between CTAS I–III and CTAS IV–V patients.
- The p-value (2.1990e-130) < 0.05, so we reject the null hypothesis, confirming that CTAS I–III and CTAS IV–V patients experience significantly different wait times.
- This means that higher-acuity patients (CTAS I–III) generally wait longer than lower-acuity patients (CTAS IV–V), reinforcing the need for optimized triage and resource allocation in ERs.

ANOVA Stastical Test

Hypothesis:

H_0 (Null Hypothesis): There is no significant difference in average ER wait times across different CTAS levels.

H_1 (Alternative Hypothesis): There is a significant difference in average ER wait times across at least one of the CTAS levels.

In [43]:
```python
# Performing ANOVA test
ctas_levels = df['CTAS_Level'].unique()
wait_time_groups = [df[df['CTAS_Level'] == level]['WaitTime_Hours'] for level in ctas_levels]

anova_result = stats.f_oneway(*wait_time_groups)

# Display ANOVA results
print("\nANOVA Test Results:\n")
print(f"F-statistic: {anova_result.statistic:.4f}")
print(f"p-value: {anova_result.pvalue:.4f}")

# Interpretation
```

```
alpha = 0.05
if anova_result.pvalue < alpha:
    print("\nConclusion: Reject Null Hypothesis H₀, concluding that at least one CTAS level has a significantly diffe
else:
    print("\nConclusion: Fail to reject Null Hypothesis H₀, meaning there is no significant difference in ER wait tim
```

```
ANOVA Test Results:

F-statistic: 2263.0866
p-value: 0.0000

Conclusion: Reject Null Hypothesis H₀, concluding that at least one CTAS level has a significantly different wait tim
e compared to others.
```

Interpretation of ANOVA Results:

- CTAS (Canadian Triage and Acuity Scale) categorizes ER patients based on urgency, from
  - CTAS I (most critical) to CTAS V (least urgent).
- The high F-statistic (2263.0866) suggests a strong variation in wait times across CTAS levels.
- The p-value (0.0000) < 0.05, so we reject the null hypothesis, confirming CTAS levels significantly impact wait times.
- This means patients with different urgency levels experience different wait times, highlighting the need for better ER resource allocation.

Chi-Square Statastical Test

<u>Hypothesis</u>:

H_0 (Null Hypothesis): There is no significant relationship between CTAS level and wait time category.

H_1 (Alternative Hypothesis): There is a significant relationship between CTAS level and wait time category.

Create a categorical column for wait time

In [44]:
```
# Define bins and labels
bins = [0, 3, 6, np.inf]  # Bins: Short (<3 hrs), Medium (3-6 hrs), Long (>6 hrs)
labels = ["Short", "Medium", "Long"]

# Create a new categorical column for wait time
df["WaitTime_Category"] = pd.cut(df["WaitTime_Hours"], bins=bins, labels=labels)
```

```python
# Display the DataFrame
print(df)
```

```
                  Hospital       Region        Date  Age  \
0            St. Michael's      Toronto  2024-12-14   55
1         Hamilton General     Hamilton  2024-03-28   27
2               Sunnybrook      Toronto  2024-10-20   74
3            St. Michael's      Toronto  2024-06-18   39
4          Toronto General      Toronto  2024-05-10   53
...                    ...          ...         ...  ...
4995      Kingston General     Kingston  2024-08-04   74
4996  Thunder Bay Regional  Thunder Bay  2024-12-30   31
4997          Ottawa Civic       Ottawa  2024-02-28   56
4998      Kingston General     Kingston  2024-11-01   49
4999  Thunder Bay Regional  Thunder Bay  2024-11-06   48

                   CTAS_Level  WaitTime_Hours  ER_Occupancy  DayOfWeek  \
0     CTAS I-III (Discharged)             2.0            84   Saturday
1      CTAS IV-V (Discharged)             1.4            97   Thursday
2     CTAS I-III (Discharged)             7.6            68     Sunday
3     CTAS I-III (Discharged)             2.8            73    Tuesday
4     CTAS I-III (Discharged)             5.2            98     Friday
...                       ...             ...           ...        ...
4995  CTAS I-III (Discharged)             7.8            10     Sunday
4996  CTAS I-III (Discharged)             1.9            83     Monday
4997  CTAS I-III (Discharged)             1.2           120  Wednesday
4998   CTAS IV-V (Discharged)             1.9           151     Friday
4999  CTAS I-III (Discharged)             0.8           103  Wednesday

      Season WaitTime_Category
0     Winter             Short
1     Spring             Short
2       Fall              Long
3     Summer             Short
4     Spring            Medium
...      ...               ...
4995  Summer              Long
4996  Winter             Short
4997  Winter             Short
4998    Fall             Short
4999    Fall             Short

[5000 rows x 10 columns]
```

Explanation:

- We created the WaitTime Category column to simplify analysis and identify patterns in ER wait times. By grouping wait times into categories (e.g., Short, Medium, Long), we can better compare trends, perform statistical tests like the Chi-Square test, and determine if factors like CTAS Level (patient urgency) influence how long patients wait.

Perform Chi-Square Test

In [45]:
```
# Create a contingency table for wait time categories vs. CTAS Level
contingency_table = pd.crosstab(df["WaitTime_Category"], df["CTAS_Level"])


# Perform Chi-Square Test
chi2_stat, p_value, dof, expected = stats.chi2_contingency(contingency_table)

# Print Results
print("\nChi-Square Test Results:\n")
print(f"Chi-Square Statistic: {chi2_stat:.4f}")
print(f"P-value: {p_value:.4f}")


# Interpret the Results
alpha = 0.05  # Significance level

if p_value < alpha:
    print("\nConclusion: Reject the null hypothesis (H₀). The distribution of CTAS levels varies across wait time cat
else:
    print("\nConclusion: Fail to reject the null hypothesis (H₀). No significant relationship between CTAS level and
```

```
Chi-Square Test Results:

Chi-Square Statistic: 2084.0424
P-value: 0.0000

Conclusion: Reject the null hypothesis (H₀). The distribution of CTAS levels varies across wait time categories.
```

Interpretation of Chi-Square Test Results:

- The Chi-Square Test evaluates whether there is a significant relationship between CTAS levels (patient urgency) and wait time categories (Short, Medium, Long).
- The high Chi-Square statistic (2084.0424) indicates a strong association between CTAS levels and wait time distribution.

- The p-value (0.0000) < 0.05, so we reject the null hypothesis (H_0), confirming that wait time categories vary significantly across different CTAS levels.
- This means that patient urgency (CTAS level) is strongly linked to how long they wait in the ER, reinforcing the importance of efficient triage systems to optimize patient flow and reduce delays.

Conclusion:

- This study analyzed ER wait times in Ontario, focusing on CTAS levels and their impact on patient flow.
- Exploratory Data Analysis (EDA) revealed that higher-acuity patients (CTAS I–III) experience longer wait times compared to lower-acuity patients (CTAS IV–V).
- Statistical tests confirmed significant differences in wait times:
  - **ANOVA results** showed that CTAS levels significantly impact ER wait durations.
  - **T-Test results** confirmed CTAS I–III patients wait significantly longer than CTAS IV–V patients.
  - **Chi-Square test** indicated a strong relationship between patient acuity and wait time categories (Short, Medium, Long).
- AI-driven triage solutions were proposed to optimize resource allocation, reduce bottlenecks, and enhance patient prioritization.
- The findings highlight the need for ER process improvements, including real-time triage monitoring, AI-based decision support, and better resource distribution to reduce wait times and improve patient outcomes.